Psychological Monographs

General and Applied

Nos. 427 and 428 1956

Graham, Matarazzo, and Caldwell

Vol. 70 Nos. 20 and 21

Behavioral Differences Between Normal and Traumatized Newborns:

I. The Test Procedures

By

Frances K. Graham

Washington University

II. Standardization, Reliability, and Validity

By

Frances K. Graham

Washington University

Ruth G. Matarazzo

Harvard University

and

Bettye M. Caldwell

Washington University

Price $1.50

Edited by Herbert S. Conrad

Published by The American Psychological Association, Inc.

Psychological Monographs:
General and Applied

Combining the Applied Psychology Monographs and the Archives of Psychology with the Psychological Monographs

Editor

HERBERT S. CONRAD

Department of Health, Education, and Welfare
Office of Education
Washington 25, D.C.

Managing Editor

LORRAINE BOUTHILET

Consulting Editors

DONALD E. BAIER	HAROLD E. JONES
FRANK A. BEACH	DONALD W. MACKINNON
ROBERT G. BERNREUTER	LORRIN A. RIGGS
WILLIAM A. BROWNELL	CARL R. ROGERS
HAROLD E. BURTT	SAUL ROSENZWEIG
JERRY W. CARTER, JR.	ROSS STAGNER
CLYDE H. COOMBS	PERCIVAL M. SYMONDS
JOHN F. DASHIELL	JOSEPH TIFFIN
EUGENIA HANFMANN	LEDYARD R TUCKER
EDNA HEIDBREDER	JOSEPH ZUBIN

MANUSCRIPTS should be sent to the Editor.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. *Background and bibliographic materials must, in general, be totally excluded* or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

The first page of the manuscript should contain the title of the paper, the author's name, and his institutional connection (or his city of residence). Acknowledgments should be kept brief, and appear as a *footnote* on the first page. No table of contents need be included. For other directions or suggestions on the preparation of manuscripts, see: CONRAD, H. S. Preparation of manuscripts for publication as monographs. *J. Psychol.*, 1948, 26, 447-459.

CORRESPONDENCE CONCERNING BUSINESS MATTERS (such as subscriptions and sales, change of address, etc.) should be addressed to the American Psychological Association, INC., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

COPYRIGHT, 1956, BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

No. 427
1956

Vol. 70
No. 20

Behavioral Differences Between Normal and Traumatized Newborns: I. The Test Procedures

By

Frances K. Graham

Washington University

Vol. 70, No. 20 Whole No. 427, 1956

Psychological Monographs: General and Applied

Behavioral Differences Between Normal and Traumatized Newborns: I. The Test Procedures[1]

FRANCES K. GRAHAM
School of Medicine, Washington University

THE present work was undertaken as a necessary steppingstone in moving toward the solution of a more complex problem. Many medical disciplines have been concerned with "birth injuries"[2]—the factors giving rise to them and their consequences, both immediate and delayed. In studying the causes of birth injury, it has become clear that they include much more than the conditions under obstetrical control, as the anesthesia, type of delivery, etc. The status of the infant himself, his size, maturity, and resistance, are of at least equal importance. Similarly, in attempting to assess the consequences of birth injury, one immediately finds grounds for the suspicion that a given external trauma will have different effects depending on such factors as whether the newborn is full term or premature, what the placental placement was, the mother's state of nutrition, etc. The number of probably significant variables that have been implicated is so large that it has seemed as if only a large-scale statistical study involving thousands of cases could possibly cope with the complexity.

The author felt that another method of attack would be necessary to answer the highly important practical question of *which* infant will show later abnormalities in development. This method was to begin with the infant rather than with external measures of trauma. If factors inherent in the individual infant introduce so much of the cumbersome complexity, then ordering infants according to the severity of their response to trauma might better predict the outcome. This is, in fact, the procedure which is employed by the physician in his clinical capacity. He makes clinical judgments or predictions about an individual infant, based on knowledge of the trauma, but based primarily upon how the infant himself responds after birth. In order to make this process scientifically useful, it is necessary to find some way of quantifying or ordering the

[1] This work was initiated by a four-month grant from the Frances Israel Fund of the Noshin Rachmonioth Society of St. Louis, Mo. From Nov. 1, 1953 to Oct. 31, 1954 it was supported by the Children's Research Foundation, St. Louis, Mo., and since Nov. 1, 1954 by research grant B685 from the National Institute of Neurological Diseases and Blindness, of the National Institutes of Health, Public Health Service. It is part of a long-term project to study the effects of anoxia, being carried out by Drs. Alexis F. Hartmann, Miriam M. Pennover, and the author, as principal investigators. Appreciation is expressed to the many others who have participated in the project, and especially to Drs. Ruth G. Matarazzo and Bettye M. Caldwell who have contributed to the psychological aspects of the research.

[2] Birth injury is defined broadly to include any condition at birth giving rise to brain injury, whether primarily from anoxia or from mechanical trauma.

clinical judgments and, preferably, of making them independent of the skill of particular practitioners.

Most important from the point of view of the consequences of birth injury, is the desirability of distinguishing the infant whose response to asphyxia is severe from one whose response is mild or absent. It has been argued on the one hand that an infant who survives the experience of anoxia at all will survive without later consequences and, on the other hand, that a large number of psychiatric and intellectual difficulties in later life may be ascribed to this cause. The literature is extensive as well as controversial but a brief summary of major studies may be found in Ford's (5, p. 951) authoritative text. It would seem that more agreement as to eventual consequences might be reached if the neonatal response were taken into account. One would expect that the later sequelae would not be the same for the infant whose condition was precarious for several days after birth as for one who appeared normal even though subjected to apparently equal trauma.

Another advantage of approaching this field by studying the infant's response is that it is possible to obtain information immediately concerning a number of problems, rather than waiting for the passage of years. Just as the mortality rates with various obstetrical procedures are of significance, so is it of interest to know how or whether the condition of a newborn infant asphyxiated for five minutes differs from one who has not been so traumatized. It is of interest to know whether use of obstetrical methods such as natural childbirth are reflected in any differences in neonatal behavior. Where no immediate differences can be found, it may be fruitless to search for them at a later stage when the number of other possibly relevant variables has increased a hundredfold.

While the desirability of considering the infant's response to trauma may be readily acknowledged, it is necessary to develop a satisfactory method of measuring newborn behavior before this variable can be introduced into experimental work. The present paper reports our efforts to do this. We have spoken of measuring "response" or "behavior," but this is an amorphous categorization which may be variously defined. For our present purpose, we are interested in measuring any response in the behavioral repertoire of the neonate which may be related to the kinds of trauma and the kinds of consequence which may be the result of birth injury. More specifically, we set as our goal the measurement of behavior which would differentiate a group of infants who were "normal" from those who might be candidates for 'brain injury." We were interested primarily in differences between normal and abnormal or traumatized infants and not in variability among normal infants.

There have been a number of systematic, psychological studies of newborn behavior but these have been restricted to normal infants. A review of these studies may be found in Pratt (13). General descriptions of the behavior of "abnormal" newborns are available in pediatric textbooks, and Gesell and Armatruda (6, p. 215 ff.) have given detailed reports of individual brain-damaged cases. This literature provided clues as to the kinds of responses which might be usefully measured, and we attempted in preliminary work to develop objective measurement techniques. Five procedures were eventually developed—measures of pain threshold, maturational level, visual response, and ratings of irritability and muscle tension. It is the purpose of this

and the following paper to report in detail the methods and their standarization and degree of differentiating power. These five tests constitute a measure of neonatal response now being emphasized in a second study as a basis for prediction of subsequent development.

Test Procedures

In selecting these particular procedures, we were guided by a number of considerations. The primary one was that the test should differentiate normal from traumatized *Ss* and that the range of possible variation should be sufficiently great so that differences would not be minimized by a low ceiling. A second consideration was the desirability of making measurement objective and reliable. We did, however, prefer to explore relatively subjective measures giving promise of validity rather than concentrate attention on developing precision of measurement in an area where the worth of measuring at all remained to be established.

It was also necessary to take into account special difficulties arising from the nature of the *Ss* and the hospital setting. In the infant, for example, the momentary internal state is a factor which may obscure the effect of other determinants of behavior if allowance is not made for it. In general, we attempted to solve this by maintaining a constant state between waking and sleeping. The specific methods of dealing with the problem are described under the several procedures.

The infant also presents special difficulty in that his repertoire of response is limited and that there seem to be few clear-cut stimulus-response connections. A great deal has already been said about generalized mass movements as characteristic of the newborn. Whether he characteristically indulges in generalized or in specialized movements, it is unquestionably difficult in many cases to determine whether movements occurring after application of a stimulus are "responses" or are spontaneous. One of our major tasks, therefore, was to discover responses which did show a specific relationship to a stimulus and to set up rules for unequivocally identifying the response.

Changes in behavior during the newborn period are very rapid, and behavior characteristic of Day 1 may not be present to the same degree on Day 2. Either the age of infants must be held constant to the day, or the relationship between age and the particular test measure must be empirically determined. In order to obtain maximum differences we wanted to examine infants on the first day. Yet the most severely traumatized *Ss*, the most likely to show marked differences, often could not be tested for several days. Holding the day of testing constant, therefore, meant either sacrificing the sensitivity gained from testing on Day 1 or sacrificing the *Ss* most severely traumatized. Rather than make either of these sacrifices, we carried out the more time-consuming method of obtaining age norms.

Other considerations more related to the nature of the project than to the nature of the *Ss* had to be taken into account. The fact that we were dealing with hospital patients, many of whom were seriously ill, necessitated the use of procedures which would be minimally disturbing to the *S*. Considerations of space, subsequent clinical use of tests, and the scope of the project disposed us to seek tests which did not take long to administer and did not require elaborate apparatus.

The most seriously limiting factor was the relatively few abnormal *Ss* available. In the well-run, modern obstetrical hos-

pital, the percentage of babies who suffer severe birth injury is only 1 to 3 per cent. This meant that an immediate check on the validity of a measure could not be made but had to wait until a sufficiently large number of traumatized *Ss* had accumulated. Probably the major effect of this was only to threaten the experimenters' morale, but it also meant that once a test procedure had been adopted and given to any appreciable number of abnormals, changes which might have improved the procedure were not made since they entailed too great a loss of difficult-to-replace data. It was because of this scarcity of traumatized *Ss* that we also included infants with mild or questionable injury rather than using more widely separated criterion groups. Except for the apparatus used in the measure of pain threshold, all materials used in this study are readily available and could easily be duplicated by others conducting research in the same area. A detailed description of the exact procedure for each technique is given in the following pages in order to facilitate such usage and to provide the necessary information for interpreting subsequent results.

Pain Threshold

A determination of pain threshold was employed for a number of reasons. Both pediatric and psychological examinations of the newborn have concentrated on motor ability and have neglected sensory functioning. Yet impairment of sensory functioning has been a useful sign of brain damage in the adult, and Windle (17, p. 57) reports that, after intrauterine asphyxiation of guinea pigs, sensory defects, when determinable, were more persistent than motor defects. We selected pain as the sensory area on which to concentrate precise measurement, after preliminary experimentation revealed various practical difficulties in the other areas. One major advantage of using pain was that a specific, discrete response could be obtained which was relatively easy to differentiate from spontaneous movement. Another consideration was the possibility that sensitivity to pain has not reached a maximum at birth. If so, pain thresholds might prove a relatively sensitive indicator of brain damage, since many lines of investigation have suggested that functions which are not well established are especially susceptible to impairment. Previous studies of sensitivity to pain in newborns have recorded any response that could be observed rather that a specific response. They have also varied the method of stimulation and have not used stimuli of graded intensity.

Our general procedure was to determine the stimulus intensity necessary to elicit a specified response within a specified time. Initially an ingenious method of mechanical stimulation developed by Bishop (2) was tried. Despite the superiority of this apparatus over previous mechanical methods, control of the rate and angle of application was found to be too difficult, at least in working with infants, and the procedure was abandoned. An electronic stimulator, also developed by Bishop (3), was substituted and found satisfactory. The apparatus delivers shocks whose intensity, duration, and frequency can be varied within the stimulating range for skin and peripheral nerves. For the present work, duration and frequency were held constant so that the shock consisted of a two-second impulse at a frequency of 14 per second which could be varied in intensity from 50 to 530 volts. Eleven voltages covering this range and corresponding to the points on a control dial were selected as stimuli. The average difference between stimuli was 50 volts. A resistance of one-half meghom was in series in the circuit, thus reducing the effect of individual skin resistance. We also made separate measurements of skin resistance in a number of *Ss* to check that, in fact, the resistance did not determine the response.

The infant was stimulated while lying on his

back with the legs in the usual flexed position. The indifferent electrode, covered with moist gauze, was placed in the center of the back against the skin, and the shirt tied so that it remained in position. A circular area, approximately ¾ inches in diameter, just below the knees of both legs was treated with vaseline. These areas were stimulated alternately by a silver ball electrode of 1 mm. diameter. An effort was made to avoid stimulating the same spot successively, especially if the skin reddened. There was apparently sufficient spread of stimulation so that differences in spot sensitivity did not introduce undue variable error.

The specific response required was a movement of the stimulated leg or the foot of that leg, with or without other bodily movement. Wickens (16) has described a similar specific response when shock was applied to the soles of the feet. While the characteristic response to the stimulus was a raising of the stimulated leg only, with foot flexion and either flexion or extension of the knee, there was considerable individual variation. Our only requirement, was, as noted above, that the stimulated limb should move. If there was any doubt as to whether or not the movement was spontaneous, as in cases with unusually short latency or where the movement appeared to be part of a generalized movement such as stretching, the stimulus was repeated. Crying or too active infants were given a pacifier but we tried to avoid stimulating at a time when the infant was sucking vigorously. Statistical analysis indicates that using pacifiers did not introduce a constant error.

There was a minimal interval of 15 seconds between presentations, longer if the infant remained active. The two-second duration of the stimulus application was timed by watching peripherally a timer with a large sweeping hand measuring to hundredths of a second. While this meant that the examiner had to observe both the response and the clock, it was not too difficult, and periodic two-observer checks on accuracy showed no consistent or appreciable error. In cases of uncertainty as to whether or not a delayed response had occurred within the time limit, the stimulus was repeated.

Four threshold determinations were made, using descending-ascending-descending-ascending order of presentation, and the average of these four series was taken as the threshold. In order to insure that the infant was sufficiently wide awake for test purposes, the first series was counted only after the elicitation of two consecutive responses. The first stimulus was one of medium intensity (315 volts, dial setting 5) to which most infants responded. If the infant did not respond to this first stimulus and to the one immediately below it in the series, stimuli of increasing intensity were given until two successive responses occurred. Occasionally when the ½-megohm resistance was left in, an infant did not respond at even the highest voltage available. In these cases, the resistance was taken out of the circuit and two shocks of 480 volts were administered. This is an increased amount of shock, but of an undetermined quantity since, with the resistance removed, the effect of individual skin resistance will vary from *S* to *S*. It was sufficiently intense, however, to elicit a response in all but two infants.

When the two successive responses had been obtained, the first descending series was continued with the stimulus one step below that to which the infant had responded. For example, if the infant responded to stimuli 5 and 4 (315 and 260 volts), the next stimulus was 3 (205 volts). If a response had not been obtained until stimuli 7 and 8 (in that order) were reached, the next stimulus was 7. When the ½-megohm resistance had been removed in order to obtain a response, it was replaced for the next stimulus and the series continued with the highest resistance-in stimulus (dial setting 10). The series was continued until the infant did not respond to two consecutive stimuli. The threshold was then calculated as the midpoint between the last stimulus evoking a response and the first stimulus which did not evoke a response.

There were two exceptions to the use of the midpoint as the threshold. Abnormal infants sometimes responded to 480 volts when the ½-megohm resistance was removed and failed to respond to dial setting 10 (530 volts) when it was replaced. As noted above, removal of the ½-megohm resistance increased the effective shock intensity to an undetermined extent so that we could not interpolate between these values. The threshold was therefore arbitrarily assigned its minimal possible value, that of 530 volts. A few older infants responded to dial setting 0 (50 volts). In these cases, a twelfth stimulus value (0 volts) was introduced by turning off the current. If the infant did not respond to the touch of the electrode alone, the threshold was, of course, the midpoint or 25 volts. If he did respond when there was no shock at all, the threshold (for response to shock) was zero volts.

The second series, an ascending one, began with the stimulus one step above that which completed the first series and continued through two consecutive responses. The third series, a descending series, was similarly begun with the stimulus one step below that which had completed the second series.

Occasionally, reversals occurred within a series, i.e., on a descending series, the *S*, after failing to respond to one stimulus, then responded to a less intense stimulus; or on an ascending series,

No reversal series (normal S)	Reversal series (normal S)	Reversal series (abnormal S)
5+	5+	5−
4+	4+	7−
→	rev.→	
3−	3−	9−
2−	2+	10−
3−	1−	480 v. +
→		
4+	0−	480 v. +
		→
5+	1−	10−
4+	2−	9−
3+	3−	10−
→	→	→
2−	4+	480 v. +
1−	5+	480 v. +
→		rev.→
2+	4+	10−
	→	
3+	3−	9+
	2−	8−
	etc.	7−
		8−
		etc.

→ Indicates a threshold.
+ Response.
− Failure-to-respond.

FIG. 1. Illustrations of the order of presenting pain stimuli.

after responding to one stimulus, he failed to respond to a more intense stimulus. Such a series was continued through the criterion of two consecutive responses or failures but was disregarded in threshold determinations. Instead additional series were presented until there were four non-reversed thresholds. Some illustrations of the presentation procedure are diagrammed in Figure 1.

Maturation Scale

A Maturation Scale was devised on the basis of items reported in previous studies (4, 6, 8, 10), and adapted to the needs of the present study and the nature of the present age-groups.[3] The primary criterion for inclusion of an item was that it should reflect changes with age, in this case differences between one-day-old prematures and one-day-old full term infants or between one-day and five-day-old full terms. A few items were included which did not differentiate well on the basis of age because preliminary work suggested that they might differentiate the traumatized infant. Other criteria used were that the scale as a whole should survey the total repertoire of behavior not covered in other tests of the battery, that the items should involve little subjectivity in scoring, that the total time for administration be about fifteen minutes, and that it not require any elaborate equipment.

The scale consists of nine items which receive varying credits depending upon the level of the response. If more than one kind of response occurs, the infant is given credit for the higher-scoring response. The general distinction between low-scoring and high-scoring responses is between generalized, mass movements and more specialized, stimulus-oriented responses. The maximum possible score is twenty-one.

Items do not have to be administered in any set order. Some items arouse the infant more completely than others, so that variation in order of presentation in such a way as to preserve a semiwakeful state was attempted. When crying occurred, testing was interrupted until the *S* was soothed with pacifier, rocking, or whatever method achieved results. The pacifier was not permitted during actual administration of the items. A record was kept of whether the *S*'s eyes were open or closed. If the infant was rated as not irritable (see below), was only minimally active and did not open his eyes on at least 50 per cent of the trials, conditions for testing were considered unsatisfactory. Test results under

[3] The author wishes to express gratitude to Ruth G. Matarazzo for her aid in making a final revision of the Maturation Scale and for her help in clarifying instructions on this and subsequent procedures.

these conditions have been analyzed separately.

One to three trials were given on each item. In order to reduce testing time, if the *maximum* credit for an item was obtained on the first trial, no further trials were given. If the *same* credit was obtained on the first two trials, no further trials were given. Otherwise, *all three trials were given and the final score for the item was the average of the credits on each trial.* Generally a second trial on an item was not given immediately after a first one, but other items were interspersed.

A trial could be discarded and another trial substituted if it was felt that test conditions were unsatisfactory due to an interruption, onset of crying, etc. However, trials were generally not discarded simply because it was difficult to determine the scoring of the behavior. The practice was followed of requiring a clear-cut response in order to receive credit. In cases of indecision, credit was not given. The record blank used in scoring this test is shown in Figure 2.

Item 1. *Prone head reaction.* The infant is placed in the prone position. His arms are extended over the head with palms down and his head is adjusted so that the nose touches the bed. He is observed for 60 seconds.

One credit is given if, *within 20 seconds,* the infant turns his head sufficiently to the side so that the nose is no longer buried.

Two credits are given if, *within 60 seconds,* the infant lifts his head so that it *just clears the bed,* whether or not he first turns it. Credit is not given for clear efforts to raise the head unless the infant succeeds in raising it sufficiently so that there is a space between his face and the bed.

Item 2. *Crawls.* This item is observed during administration of Item 1.

One credit is given *if at least three alternating movements of the leg,* simulating crawling, occur in succession. Flexion of both legs followed by extension of both legs is not credited.

Item 3. *Pushes feet.* With the infant lying on his back, the examiner places one hand flat against the bottom of both feet in such a way that the infant's knees are flexed. Gentle pressure is exerted against the feet for 20 seconds.

One credit is given if there is a definite increase in the muscle tone of the legs such that the examiner can feel the increased pressure against his hand and can observe the tightening of the leg muscles.

Two credits are given if the infant extends his legs and thus pushes the examiner's hand *or,* with the legs remaining flexed, pushes against the hand with sufficient force to move himself backward. Occasionally, a Moro or a general stretching movement occurs during the administration of this item. The trial should then be repeated.

Item 4. *Auditory reaction.* With the infant lying on his back, a rattle is shaken gently several times at a distance of about six inches from either side of the head and in line with the ear. The rattle should not be in the infant's field of vision. Examiner shakes the rattle for approximately 5 seconds and then pauses for 5 seconds. This procedure is repeated for four presentations of the rattle, but with each presentation the intensity of the shaking is increased. The range is roughly from a very slight, minimal rattle to as abrupt and sharp a rattle as can be given without at the same time shaking the table or infant. After a pause of 10 to 20 seconds, a bell is sounded. This procedure was adopted because the increasing intensity of stimulation normally elicited more than one kind of behavior and made it easier to distinguish the different kinds of response scored. While more precise control of the stimulus might be desirable, it was felt that the present procedure satisfied the aims set up for the Maturation Scale. Not only would precise control of the auditory stimulus require elaborate apparatus, but there is perhaps more likelihood of eliciting several kinds of response when the quality of the sound is varied by using both rattle and bell and using noise rather than pure tones.

One credit is given for any change in behavior which is clearly in response to the bell, i.e., which occurs within 2 seconds after the bell is sounded. A Moro or partial Moro is a common response. Eye blinking or tightening of the eye muscles is a common minimal response and is given credit.

Two credits are given if the infant responds to the rattle. The same criteria are applied as for the bell.

Three credits are given if the infant appears to listen to the rattle. Opening of the eyes and decreased movement or, in older infants, head turning and eye movements, whether or not in the direction of the stimulus, are credited. This should not be scored unless it is a clear-cut response.

	Trial	1	2	3
1.	Head Reaction (prone position) 60"			
	1 Turns head to side in 20".	--	--	--
	2 Head clears bed in 60".	--	--	--
2.	Crawls (prone position) 60"			
	1 Three alternating leg movements simulating crawling.	--	--	--
3.	Pushes feet (supine) -- gentle pressure exerted for 20".			
	1 Increase in muscle tone of legs.	--	--	--
	2 Legs extend, or body moves backward.	--	--	--
4.	Auditory Reaction (supine) -- 5" S periods.			
	1 Response to bell (startle, blink).	--	--	--
	2 Response to rattle (as above).	--	--	--
	3 Listening to rattle (opening eyes, decreased movement, head turning).	--	--	--
	4 Persistent listening to rattle (3 of 4 presentations), or listening also to bell.	--	--	--
	5 Both persistent listening to rattle and listening to bell.	--	--	--
5.	Cotton (supine) 20"			
	1 Any movement within 2".	--	--	--
	2 Specific movements of either head or mouth (back and forth movements or sustained head retraction or movements of mouth in direction of S).	--	--	--
	3 Both specific head and mouth movements present more than 50% of R time.	--	--	--
	4 Head, mouth, and coordinated arm movements all present.	--	--	--
6.	Paper (supine) 20"			
	1 Same as Item 5.	--	--	--
	2 Same as Item 5.	--	--	--
	3 Same as Item 5.	--	--	--
	4 Same as Item 5.	--	--	--
7.	Persistence -- percentage of time infant responds to S in Items 5 and 6.			
	1 Persistent R in 2/3 of trials.	--	--	--
	1/2 Persistent R in 1/2 of trials.	--	--	--
8.	Vigor -- rate and extent of movement in 5 and 6.			
	1 Vigorous R in 2/3 of trials.	--	--	--
	1/2 Vigorous R in 1/2 of trials.	--	--	--
9.	Grasp (supine).			
	1 Infant registers pull of more than 30 ounces.	--	--	--

TOTAL SCORE: _____

FIG. 2. Maturation Scale record blank.

Four credits are given if either *(a)* the listening to the rattle is persistent through three of the four presentations of the rattle or *(b)* if the listening response occurs only once to the rattle but also occurs to the bell, even if followed after a brief quieting by a typical Moro response.

Five credits are given if both persistent listening to the rattle and listening to the bell occur.

Item 5. *Cotton.* With the infant lying on his back, a small piece of cotton is lightly placed so that it covers the nostrils and barely touches the upper lip. It is held in position for 20 seconds.

One credit is given for any change in behavior which is clearly a response to the stimulus. Any easily discriminated movement is credited if it follows within two seconds the application of the stimulus. Responses with a longer latency are credited if the examiner feels they are not spontaneous. Long latencies are more likely to be found in relatively inactive infants and therefore, if a fairly vigorous response involving the extremities and at least one head turn occur at any time during the 20-second presentation period (and there is no other stimulation), it may be credited. If only small movements or twitches occur, they are generally spontaneous.

Two credits are given if the response includes specific movements, of *either the head or mouth*. Credit for specific head movement is given for (*a*) several back and forth movements or (*b*) a sustained head retraction which is usually a coordinated body movement of avoidance and involves arching of the back. Specific mouth movement is scored when there is sustained opening of the mouth with movements of the lips upward in the direction of the stimulus. Sucking, yawning, or small lip movements with the mouth closed are not credited. To receive credit at this level, the specific movements should be present a majority of the time that any response is present, i.e., the characteristic mode of responding should include specific oriented movements. If the typical response is a generalized mass movement with an occasional specific response, credit is not given at this level.

Three credits are given if *both* specific head and mouth movements are present more than 50 per cent of the time the infant is responding. Criteria for a specific response are the same as above.

Four credits are given if head, mouth, and coordinated arm movements are all present. Coordinated arm movements are scored if both hands are brought to the midline more than once or are kept at the midline, and if the hands are open. The infant in these cases appears to be batting at the stimulus. A single such coordinated movement is not credited, but there may be some alternation in the kind of specific response, as long as the characteristic mode of responding is with specific oriented movements. Coordinated arm movements without head and mouth movements did not occur in our sample. Had they occurred, three credits would have been given.

Item 6. *Paper*. With the infant lying on his back, a piece of cellophane, 1½" by 2", is lightly held with one hand in such a way that it covers both the nostrils and the mouth. It is held in position for 20 seconds.

Zero to 4 credits are assigned in the same manner as for Item 5.

Item 7. *Persistence*. During each of the trials for Items 5 and 6, the examiner notes the percentage of time during which the infant is responding to the stimulus. As we found it somewhat difficult to observe the response closely and also follow a stop watch, this judgment is not a precise one. The infant is credited as persistent on a trial if he responds throughout the 20-second period. If there is a period as long as 3 or 4 seconds during which he does not respond, whether initially, during, or near the end of stimulation, persistence is not credited. The number of trials on which this judgment is made will vary from two to six, depending on how many are necessary to score Items 5 and 6. Persistence is not scored as either present or absent on a trial when there is no response to the stimulus. However, if there is no response on all 6 trials, the persistence credit is, of course, zero.

One credit is given if a persistent response occurs in two-thirds of the trials.

One-half credit is given if a persistent response is obtained in one-half of the trials, *or* if the examiner is in doubt as to whether the behavior on the majority of the individual trials should or should not be called persistent. Half-credits are allowed on this item since it requires a more subjective rating than most.

Item 8. *Vigor*. The procedure for this item is the same as for Item 7. This also requires a relatively subjective rating and half-credits are therefore permitted. The criterion of vigor was the behavior of the normal, full-term, one-day-old infant. Prematures were generally given half-credit on this item. Rate and extent of movement primarily determine the judgment of vigor, and duration the judgment of persistence. Persistent and vigorous behavior tend to occur together but may occur independently.

One credit is given if a vigorous response occurs in at least two-thirds of the trials.

One-half credit is given if a vigorous response occurs in one-half of the trials *or* if the examiner is in doubt as to whether the behavior on the majority of the individual trials should or should not be called vigorous.

Item 9. *Grasp*. With the infant lying on his back, a rubber-covered metal stirrup attached to a small spring balance is placed in the infant's palm. The apparatus and procedure are adapted from Sherman *et al.* (14). The stirrup is rubbed against the palm until the infant grasps it. A gentle, steady pull is then exerted until the infant either releases his grasp or the scale registers a pull greater than 30 ounces.

One credit is given if the strength of pull reaches or exceeds 30 ounces.

Vision Scale

Clinical eye examinations of the newborn are generally not extensive, although abnormalities such as ptosis, strabismus, nystagmus, immobility, and randomness of eye movements may be recorded when they are striking. Because these visual abnormalities are reported to occur frequently in normal newborns (5, p. 958; 12, p. 77), they are not considered of much value in diagnosis. No effort is ordinarily made to observe the eye's response to a stimulus—that is, to note the ability to fixate or pursue a bright light or similar object—as widespread medical opinion holds that these capacities are not present until several weeks after birth.

It has been reported, however, that the newborn, and even the late-stage fetus, can fixate briefly and can pursue a stimulus through an arc of 30 degrees (1; 7, p. 76 ff.). We were able to corroborate this readily. It is, of course, true that the newborn differs markedly in these respects from the two-month old. No special effort is required to elicit visual response from the older infant. His facial expression and the smoothness of his eye movements are sufficiently like those of an adult so that even the casual observer feels convinced that the infant "sees." With the newborn, on the other hand, patience and careful procedure are necessary. In the first place, it is difficult to obtain a satisfactory observation period when the child is neither asleep nor crying. Further, fixation and pursuit, when they do occur, are less clearly defined and of shorter duration. Thus, careful objective criteria for judging the response must be specified.

Despite these obstacles, it seemed worthwhile to include observation of visual performance in our test battery. A number of writers (7, p. 248; 15, p. 168) have felt that reading problems or defects in other complex visual perceptions may occur as a result of brain injuries sustained at birth. Certainly visual impairment is a common consequence of brain injury and probably one which is frequently undetected. In many cases, there is no structural change observable in the peripheral organ, and the impairment is demonstrated only through use of sensitive tests. Psychologists have found, for example, that brain-damaged individuals with normal vision may be less able to fuse flickering lights (11) or may have difficulty with such a complex visual-motor task as copying designs (9). It is possible that brain injuries at birth may not only produce undetected visual dysfunction but that this may more readily occur precisely because visual functions are in the process of developing and are therefore more susceptible to injury.

The general procedure we followed, that of observing and categorizing the response to a moving stimulus, is similar to the procedure used by Gesell (7) and adopted by other designers of infant tests.

Observations can be made only when the infant's eyes are open for a sufficient length of time. With most *Ss* this was not a major problem. They were under examination for half an hour to an hour, and during this time usually opened their eyes on several occasions. Other procedures were interrupted to take advantage of the opportunity, and the observations were repeated if the initial observation period had been too short or was otherwise unsatisfactory. If an *S* did not spontaneously open his eyes, we attempted to arouse him by various kinds of manipulation, of which the most successful was to hold him vertically

either upright or upside down. If it was impossible to obtain a satisfactory observation period, the test could not be included. An observation was considered satisfactory if at least five trials had been given or an *S* had obtained a score of six in fewer trials.

The stimulus is presented with the infant in a supine position. Either the examiner's hand, a bell, or a metal tape measure 1 and 3/8 inches in diameter was used. No further effort was made to standardize the stimulus, since we were attempting to determine whether the capacity was present under the most favorable conditions. As some infants seemed to respond more readily to one stimulus than to another, the examiner used whichever one was preferred. Stimuli were generally changed after one or two trials, as there appeared to be a change in their effectiveness after repeated use.

The newborn usually does not turn head and eyes toward a stimulus at the periphery. Fixation is limited to a relatively narrow range directly in front of the eyes, an area which is determined by the tonic-neck reflex. Stimuli are therefore initially presented in this area. The experimenter observes the direction in which one or both eyes appear to be turned and places the stimulus in line with the eyes. He then moves it slowly toward and away from the infant, since the distance from the eyes at which an object can be fixated is also limited, varying from about 3 inches to 12 inches (7, p. 191). It may be necessary to shift the stimulus laterally in order to locate the place of regard. If fixation does not occur, the examiner attempts to elicit it by adding an auditory stimulus temporarily (e.g., the bell may be rung softly, the fingers snapped, etc.).

After preliminary efforts to locate the position which is optimal for fixation, there are presented a number of trials in which the stimulus is moved slowly upward from the line of regard or in either horizontal direction for as great a distance as the infant's eyes will follow. Each trial is begun with the stimulus in the place of regard. Visual items are scored both for the type of response which can be elicited and for ease of elicitation. As many trials may be given as are necessary to arrive at these judgments. Undoubtedly this procedure introduces a subjective element. However, it appears to be more adequate than specifying an arbitrary number of trials, since so many difficult-to-define factors may make one trial unsatisfactory—the infant may grow sleepy, may be startled by an extraneous noise, may fixate on something else, may begin to hiccup, etc.

Several aspects of the visual response are considered in making a classification: (*a*) presence or absence of a kind of response (such as fixation or pursuit); (*b*) the ease of eliciting the response; (*c*) the direction of eye movement (horizontal or vertical); (*d*) the distance the eye moves. These characteristics of the response were ordered according to difficulty in such a way that success on one item, such as horizontal pursuit through 30 degrees with difficulty, in most cases automatically credited the *S* with success on lower items. Likewise, failure on one item meant failure on higher items. The ordering is partly on an a priori basis—i.e., an *S* credited with "easily-elicited fixation response" is assumed to have passed "difficult-to-elicit fixation response." Other items, as whether "short vertical pursuit" is easier or more difficult than "horizontal pursuit through 90 degrees" were ordered on the basis

of the preliminary findings, but credit was given for each only when the response could be elicited. Each response (or item), with the exception of the zero point, is given an arbitrary weight of one, so that the *number of items passed* is the *S*'s score.

Credit 0. There is no response of fixation or pursuit *and* one of the following abnormal features is observed: (*a*) wandering, uncoordinated movements (the eyes move constantly and generally independently); (*b*) immobilization or staring; (*c*) pinpoint pupils, although the eyes are spontaneously opened and remain open for more than a few seconds. The zero score is given only when these abnormal features are the characteristic response during observation rather than an isolated episode of nystagmus, strabismus, etc.

Item 1. No fixation or pursuit but none of the above abnormalities is observed.

Item 2. Fixation is brief or obtained with difficulty. The criteria of fixation have been summarized by Gesell (7, p. 191) as follows: "Quieting of body activity and of fussing; raising of upper eyelids and eyebrows; widening of eye slits, wrinkling of forehead, alignment of eye or eyes with stimulus, and immobilization of one eye or both." The criteria which we considered essential were those relating to the position of the eye or eyes, i.e., alignment with the stimulus and immobilization. Various other postural accompaniments, as noted above, generally occur and probably contribute to the certainty of the subjective judgment.

Item 3. Fixation is clearly present and easily elicited, i.e., occurs in a majority of the trials.

Item 4. Horizontal pursuit is questionably present. Pursuit is considered to be present when one or both eyes, having fixated a stimulus, move in the same direction as the moving stimulus. If the reaction time is short and the infant's eye follows the stimulus relatively smoothly and moves at about the same rate as the stimulus, there is little difficulty in identifying this as a pursuit movement. In some cases, however, repeated trials elicit delayed eye movements in the direction of the stimulus which do not appear to be random movements but which are not clearly pursuit movement. The present classification includes such cases.

Item 5. Horizontal pursuit movements through an arc of 30 degrees are obtained with difficulty. This classification does not indicate uncertainty as to whether or not there is a pursuit movement, but rather whether the movement could be elicited in fewer than 50 per cent of the trials.

Item 6. Horizontal pursuit movements through an arc of 30 degrees are easily obtained, that is, are present in more than 50 per cent of the trials.

Item 7. Vertical pursuit movements are obtained. These may be very short, but there is usually little problem in deciding whether they are present. Those movements of the eye which are upward random deviations generally begin with the eye in a peripheral position from which it rolls slowly upward. They are readily distinguished from the rapid vertical movements which follow fixation and may occur in pursuit of the stimulus.

Item 8. Horizontal pursuit movements through an arc of 90 degrees are obtained.

Item 9. Vertical pursuit is obtained through an arc of more than 30 degrees.

Item 10. Horizontal pursuit movements through an arc of more than 90 degrees are obtained. To accomplish this, the infant must move the head as well as the eyes.

Irritability and Muscle-Tension Ratings

In pediatric discussions of the brain-injured newborn, there are frequent references to such symptoms as a high-pitched and feeble cry, excessive irritability, muscular rigidity or flaccidity, and poor muscle tonus. We attempted to provide a crude quantification of these characteristics by means of two ratings —a rating of irritability and one of muscle tension or tone. The ratings are based on observations of the infant while other tests are being administered, and on a few simple supplementary procedures.

The experimenters had had considerable experience with the phenomena to be rated, and the ratings probably depend to a considerable extent upon "expertness." However, they have several advantages over the alternative of using reports in a medical chart which have been contributed by many different people. In addition to the immediate advantage of providing ordinal measurement rather than verbal description, finer discriminations become possible. Checks with the pediatric record showed that we

did not overlook cases described in these terms but that we did detect less extreme examples.

In an attempt both to reduce the dependence on "expert judgment" and to increase reliability, we analyzed the behavior upon which such ratings would be based and broke it down into separate components. The *Ss* were first rated in terms of these separate components or submeasures, and only afterward was an over-all rating made on that scale.

Irritability Rating. Irritability as used here is similar to the concept of physiological irritability—how sensitive the infant is to stimulation and not how loudly and lustily he cries. Several factors which contribute to this rating could be abstracted:

1. *The intensity of stimuli which evoke crying.* The normal infant does not usually cry when he is picked up and brought to the examining room and will be disturbed only by the more intense forms of stimulation—electrical stimulation above the threshold, loud noises, rough handling, etc. Abnormal infants, on the other hand, may squeal when first picked up and continue to fuss whenever handled or stimulated.

2. *The state of the infant.* When a normal infant cries, it appears to be because he has become too wide awake, is hungry, or is defecating. Hunger may be suspected in an infant who actively roots if the mouth is brought into contact with any surface and who vigorously sucks the offered pacifier. The normal infant also does not sustain well a too-long wakeful period. If the infant is awake when brought to examination or wakes up almost at once, he generally remains contented only for a brief period and then fairly abruptly bursts into lusty crying. In contrast, the abnormal infant, although fussing almost continuously, may remain with eyes closed.

3. *The cry of the infant.* The normal infant cries loudly and lustily. A weak, high-pitched, "fussy" cry or whimper is characteristic of the abnormal child.

4. *Ease of quieting.* The cry of the normal infant may be sustained, and he will not quiet spontaneously—suggesting that internal, not external, stimuli are responsible for the crying. The cry of the abnormal infant, on the other hand, is more likely to be intermittent and will usually cease temporarily if the infant is left alone. Occasional abnormal babies are irritable throughout but are distinguished by the characteristics noted above.

A three-point scale, with numerical values of 0, 1, and 2, was provided for the rating of irritability. It was decided in advance that the zero point should represent the behavior of a normal infant, the endpoint (value 2.0) the extreme and easily identified form of behavior seen in a grossly abnormal infant, and the midpoint (value 1.0) a "just-perceptible" form of abnormal behavior. With these points as signposts, *Ss* were actually located anywhere on the continuum. Each of the points on the scale was anchored as follows:

Normal irritability. Crying, if it occurs at all, is only in response to fairly intense external stimulation or in response to fatigue, hunger, or defecation. The cry itself is loud and may be sustained.

Just-perceptible irritability. Crying or "fussing" occurs in response to mild stimulation and does not appear to be determined by normal internal states of discomfort, as described under 2. Each episode of crying may be brief, but as much as half of the period is so occupied.

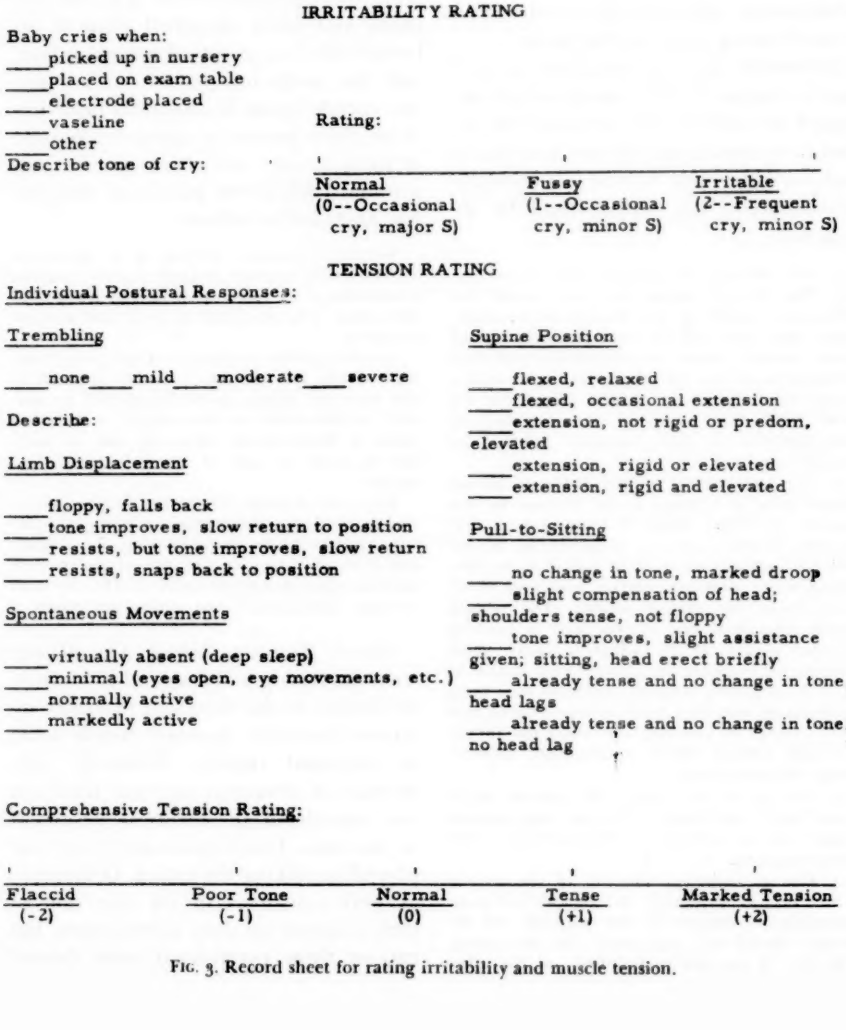
Abnormal irritability. Crying occurs in response to many mild stimuli. While the infant appears to be uncomfortable, the discomfort does not seem due to any of the normal internal states described under 2. The cry itself is weak, high-pitched, and almost continuous.

Muscle Tension Rating. The Muscle Tension Rating was designed to measure deviations in the direction of either increased flaccidity (lessened muscle tone) or increased rigidity. However, only amount of deviation and not direction was considered in statistical treatment of the data. Five submeasures were employed in making the rating. Ordinarily, *Ss* were consistent in the direction of their response on these submeasures, but two or three exceptional cases showed

mild deviations in both directions. They were rated accordingly and the deviations totaled without regard to sign.

The five submeasures required rating the following: 1, nature (flexed or extended) of the supine position which the legs assumed spontaneously; 2, resistance to limb displacement; 3, change in muscle tone in response to being pulled to a sitting posture; 4, amount of spontaneous activity; and 5, frequency of trembling of body parts and the stimuli evoking this response. The procedure for obtaining these was as follows:

1. *Leg position.* During the threshold procedure, examiner observes the leg position which was spontaneously assumed between trials. The record sheet lists five descriptive classifications of flexion-extension arranged to form a con-



IRRITABILITY RATING

Baby cries when:
____picked up in nursery
____placed on exam table
____electrode placed
____vaseline
____other
Describe tone of cry:

Rating:

Normal	Fussy	Irritable
(0--Occasional cry, major S)	(1--Occasional cry, minor S)	(2--Frequent cry, minor S)

TENSION RATING

Individual Postural Responses:

Trembling

____none ____mild ____moderate ____severe

Describe:

Limb Displacement

____floppy, falls back
____tone improves, slow return to position
____resists, but tone improves, slow return
____resists, snaps back to position

Spontaneous Movements

____virtually absent (deep sleep)
____minimal (eyes open, eye movements, etc.)
____normally active
____markedly active

Supine Position

____flexed, relaxed
____flexed, occasional extension
____extension, not rigid or predom, elevated
____extension, rigid or elevated
____extension, rigid and elevated

Pull-to-Sitting

____no change in tone, marked droop
____slight compensation of head; shoulders tense, not floppy
____tone improves, slight assistance given; sitting, head erect briefly
____already tense and no change in tone; head lags
____already tense and no change in tone; no head lag

Comprehensive Tension Rating:

Flaccid	Poor Tone	Normal	Tense	Marked Tension
(-2)	(-1)	(0)	(+1)	(+2)

FIG. 3. Record sheet for rating irritability and muscle tension.

tinuum (see Fig. 3). After the pain threshold had been determined, a check was placed after the classification which described the position or positions assumed frequently.

2. *Resistance to limb displacement.* The examiner places his hand under the infant's heel, raises the leg through an arc of 45 degrees, and then releases it. Examiner attends both to the change in tone while lifting the leg and to the response of the leg when released. This procedure is repeated several times at intervals. The response is rated by placing a check after one of the four descriptions on the record sheet.

3. *Pull-to-sitting.* The examiner takes hold of both of the infant's hands and pulls gently two or three times without raising the child from the table, noting whether the arms respond with increased muscular tonus. Examiner then grasps the infant's forearms and pulls the body forward about 45 degrees, noting the head and shoulder response. The procedure is repeated *two or more times* before the response is rated.

4. *Spontaneous movement.* This rating is made on the basis of behavior during the entire examination. A check was placed after one of the four descriptions on the data sheet.

5. *Trembling.* Trembling of body parts is observed during the whole period of examination. A qualitative description written of such movements includes the body parts involved in such a response; how frequently it occurs; and whether it is spontaneous, in response to mild stimuli, or in response to stimuli also evoking the Moro reflex.

At the end of examination, the overall rating on muscle tension was made. A five-point scale, with numerical values from -2 to $+2$, was designed for the rating. The zero point represented the behavior of a normal infant and the endpoints (values -2 and $+2$) the extremes of flaccidity and rigidity, respectively, as seen in a grossly abnormal infant. The -1 and $+1$ points represented a "just perceptible" form of abnormal behavior. Each of the points was described in terms of the submeasures, as follows:

Abnormal rigidity. 1, Characteristic leg position is extended, rigid and elevated. 2, The legs resist displacement and snap back to the initial position when released. 3, The muscles are already tense and do not change tone in response to pull-to-sitting. There is no head lag. 4, There may be any amount of spontaneous activity. 5, Trembling of more than one body part is frequent and may be spontaneous or in response to mild stimuli as well as to startle stimuli.

Just perceptible rigidity. 1, Spontaneous leg position is extended but not rigid or elevated. 2, The legs resist displacement and snap back to the initial position when released. 3, The muscles are already tense and do not change tone in response to pull-to-sitting. There is, however, head lag. 4, There may be any amount of spontaneous activity. 5, There is usually some trembling involving more than one body part, which is pronounced in response to startle stimuli and less intense in response to other milder stimuli.

Normal muscle tone. 1, The spontaneous supine leg position is flexed, or flexed with occasional extension. 2, Muscle tone improves in response to limb displacement, although there may also be slight resistance. The limb returns relatively slowly to its initial position. 3, Muscle tone of arms and shoulders improves in response to pull-to-sitting. There is usually resistance to the pull and some head lag, but an occasional superior normal does not show head lag. These Ss must be distinguished from the tense infant in whom absence of head lag is not the result of improving muscle tone but of tension in the whole body so that he moves in one piece without the uneven distribution of tone characteristic of the normal newborn. 4, There may be any amount of spontaneous activity except the condition of minimal activity with eyes open. 5, Trembling, if it occurs, involves only one or two body parts, is not extensive or of long duration, and with an occasional exception, is only in response to startle stimuli.

Just perceptible flaccidity. 1, The supine leg position is always flexed. 2, There is no change in tone in response to limb displacement. The leg falls back to its initial position when released. 3, There is no change in tone in response to pull-to-sitting and there is head lag. 4, Spontaneous activity is minimal. The eyes may be open or closed. 5, There are no more than one or two instances of mild trembling.

Abnormal flaccidity. 1, Supine leg position is always flexed. 2, The initial tone of the leg is poor, feels "floppy" and weightless. There is no change in tone in response to limb displacement, and the leg falls back to its initial position when released. 3, The arm and shoulder muscle tone is initially poor and remains so even in response to pull-to-sitting. The head droops markedly. 4, Spontaneous activity is minimal even though the eyes are open. 5, There is no trembling.

SUMMARY

Five test procedures developed to provide a means of differentiating normal newborns from those who have been traumatized and possibly brain injured are here described. These were developed in the hope that differences in initial response to trauma might aid in predicting the course of subsequent development. The five tests consist of a Pain Threshold Test, a Maturation Scale, a Vision Scale, an Irritability Rating and a Muscle Tension Rating. Standardization, reliability, and differentiating ability of the procedures are presented in the second section of the study, Standardization, Reliability, and Validity (*Psychol. Monogr.*, 70, No. 21, Whole No. 428.)

REFERENCES

1. BEASLEY, W. C. An investigation of related problems in the vision of new-born infants. *Psychol. Bull.*, 1933, **30**, 626. (Abstract)
2. BISHOP, G. H. Relation of pain sensory threshold to form of mechanical stimulator. *J. Neurophysiol.*, 1949, **12**, 51-57.
3. BISHOP, G. H. A simple electronic stimulator suitable for peripheral nerve and skin sensory testing. *EEG clin. Neurophysiol.*, 1953, **5**, 105-106.
4. CATTELL, PSYCHE. *The measurement of intelligence of infants and young children.* New York: Psychological Corp., 1940.
5. FORD, F. R. *Diseases of the nervous system in infancy, childhood and adolescence.* (3rd Ed.) Springfield, Ill.: Charles C Thomas, 1952.
6. GESELL, A., & ARMATRUDA, CATHERINE. *Developmental diagnosis.* New York: Hoeber, 1941.
7. GESELL, A., ILG, FRANCES L., & BULLIS, G. D. *Vision, its development in infant and child.* New York: Hoeber, 1949.
8. GILLILAND, A. R. *The Northwestern Intelligence Tests.* New York: Houghton Mifflin, 1951.
9. GRAHAM, FRANCES K., & KENDALL, BARBARA S. Performance of brain-damaged cases on a memory-for-designs test. *J. abnorm. soc. Psychol.*, 1946, **41**, 303-314.
10. GRIFFITHS, RUTH. *The abilities of babies.* New York: McGraw-Hill, 1954.
11. HALSTEAD, W. C. A power factor (P) in general intelligence: the effect of brain injuries. *J. Psychol.*, 1945, **20**, 57-64.
12. PARMALEE, A. H. *Management of the newborn.* Chicago: Year Book Publishers, 1952.
13. PRATT, K. C. The neonate. In L. Carmichael (Ed.), *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954. Pp. 215-291.
14. SHERMAN, M., SHERMAN, IRENE, & FLORY, C. D. Infant behavior. *Comp. Psychol. Monogr.*, 1936, **12**, No. 4.
15. STRAUSS, A. A., & LEHTINEN, LAURA E. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1947.
16. WICKENS, D. D., & WICKENS, CAROL. A study of conditioning in the neonate. *J. exp. Psychol.*, 1940, **26**, 94-102.
17. WINDLE, W. F. *Asphyxia neonatorum. Its relation to the fetal blood, circulation and respiration and its effects upon the brain.* Springfield, Ill.: Charles C Thomas, 1950.

(Accepted for publication December 19, 1955)

No. 428
1956

Vol. 70
No. 21

Behavioral Differences Between Normal and Traumatized Newborns: II. Standardization, Reliability, and Validity

By

Frances K. Graham
Washington University

Ruth G. Matarazzo
Harvard University

and

Bettye M. Caldwell
Washington University

Vol. 70, No. 21 Whole No. 428, 1956

Psychological Monographs: General and Applied

Behavioral Differences Between Normal and Traumatized Newborns: II. Standardization, Reliability, and Validity[1]

FRANCES K. GRAHAM
School of Medicine, Washington University

RUTH G. MATARAZZO
School of Medicine, Harvard University

AND BETTYE M. CALDWELL
School of Medicine, Washington University

IN AN effort to measure behavior which would differentiate a group of infants who were "normal" from those who might be diagnosable as cases of "brain injury," the five test procedures described in the preceding paper (7) were developed. The effectiveness of these procedures is reported in the present paper.

SUBJECTS

Our subjects (*Ss*) were full term infants born on the inpatient service of the St. Louis Maternity Hospital during the period from July, 1953 through October, 1955.[2] On the basis of a sample from a private, teaching hospital, it is not possible to generalize about the frequency with which certain kinds of abnormality will occur in the population as a whole. It is probable that both socioeconomic level and techniques of prenatal and postnatal care reduced the amount of abnormality in our group. This is not a serious limitation in interpreting our results, however, since we are not concerned with absolute frequencies but with relative differences between normal and traumatized *Ss*. Whatever differences we could demonstrate between normal and traumatized groups would undoubtedly be more striking in the population as a whole where abnormality would be more marked.

The traumatized group was composed of almost the total population of traumatized infants born during the course of the study, with the exclusion of infants who were overlooked, and of a few infants whom we could not obtain permission to examine. We were informed

[1] This investigation is part of a long-term project being carried out in collaboration with Drs. Alexis F. Hartmann and Miriam M. Pennoyer. The work was initiated by a four-month grant from the Frances Israel Fund of the Noshin Rachmonioth Society of St. Louis, Mo. From Nov. 1, 1953 to Oct. 31, 1954 it was supported by the Children's Research Foundation, St. Louis, Mo., and since Nov. 1, 1954 by a research grant B685 from the National Institute of Neurological Diseases and Blindness, of the National Institutes of Health, Public Health Service.

[2] The writers are indebted to the authorities of the St. Louis Maternity Hospital and to the obstetricians and pediatricians who provided the opportunity for carrying out the study. We should like to express special gratitude for the cooperation of the nursing staff and especially of Miss Margaret Weber.

TABLE I

CLASSIFICATION OF TRAUMATIZED SUBJECTS ACCORDING TO KIND AND DEGREE OF TRAUMA

Kind of Trauma	Degree of Trauma		
	Mild	Moderate	Severe
Anoxia	21	26	11
Mechanical trauma	0	0	3
Infections or diseases*	3	12	5
Total *N*	24	38	19

* Erythroblastosis fetalis, hypoglycemia, meningitis.

by the pediatric staff of all infants who, either at birth or subsequently, might be classified as abnormal. Infants were tested, if possible, within 24 hours after birth or as soon after that as their condition permitted. Only infants seven days old or younger are included in the study.

The kinds and degree of trauma present in the abnormal group are shown in Table 1. A pediatrician experienced in the neonatal field made the classifications without knowledge of psychological test results. A second pediatrician also rated 25 of the same cases.[3] Pediatrician One consistently rated cases more severely than Pediatrician Two, but the extent of agreement was high and significant as estimated by a correlation ratio of .86. It is important to emphasize that there is no reason to expect that all traumatized *Ss* would sustain actual brain damage and that those classified as "mild" are considered by both pediatric judges to have good prognoses.

The normal group was composed of infants without prenatal, perinatal, or postnatal complications. Cases were not included where there was maternal bleeding during pregnancy or serious maternal illnesses such as rubella, diabetes, hypertension, etc. Perinatal circumstances were considered satisfactory when delivery was spontaneous or by low forceps, respiration and cry were established in a few seconds, and the infant was active. Infants were not tested on the day of circumcision or with elevated temperatures. Sufficient numbers were tested on each of the first five days of life so that the effect of age could be either statistically weighted or controlled by pairing normal with traumatized *Ss*.

The primary requisite in selecting a normal control group is that it shall not differ from the experimental group except in regard to the independent variable. Unfortunately, when the independent variable is not under the control of the experimenter, there is danger that characteristics inherent in the *S* may be associated with both the independent and dependent variables and thus confound the results. There is no certain way of avoiding this, and the best that can be done is to control for such characteristics as experience indicates may be significant. In this instance, there is reason to think that prematurity, sex, race, and socioeconomic status are associated with the occurrence of trauma. However, unless they are also related to the dependent variable, i.e. test performance, the results will not be confounded and there will be no advantage to matching or pairing *Ss* or otherwise controlling for possible effects.

Preliminary results indicated that,

[3] We are grateful to Drs Miriam Pennoyer and Marshall Greenman, of the Newborn and Premature Service of the St. Louis Maternity Hospital, for assistance in this phase of the study. Originally, they were asked to group the cases into as many discriminable ranks as possible, but their judgments were uncertain when more than three ranks were attempted. These three ranks were roughly identified as mild, moderate, and severe trauma or good, questionable, and guarded prognosis. The latter description is probably preferable and has been adopted in subsequent work.

with the exception of prematurity, these characteristics were *not* related to most of our tests. We therefore decided to test full term normal infants, without any predetermined selection criteria other than those determining classification as normal, and then to test statistically for relationships between *S* characteristics and test performance and to control such relationships as were found to exist.

The exact procedure used in selecting *Ss* was determined by the supply available in the nursery on any given day. Because of the scarcity of traumatized *Ss*, a priority system was set up. Any traumatized *Ss* available were examined first. No selection was involved here, as all such *Ss* with the exceptions noted above were examined. Next, all infants on whom blood oxygen saturation tests had been made were examined. At intervals during the period of study, blood samples were obtained from all infants born during the hours the research pediatrician was on duty. This was an unselected group of mixed normals and abnormals. All of these infants were tested at least for their pain thresholds, but the results of prematures, infants delivered without difficulty by section, and otherwise normal infants whose mothers had accidents or serious illnesses during pregnancy are not included in the present report. They do not meet the criteria for classification as either normal or traumatized. Finally, if time permitted, normal *Ss* were selected from a survey of the hospital charts. It had been planned originally to obtain all *Ss* for the normal standardization group from those on whom oxygen tests had been made. However, as there were a number of interruptions in carrying out the oxygen-testing program, it became necessary to find additional *Ss*. The psychologist looked through the hospital charts and selected the first infant who met the criteria for normality. Negro and white patients and, to some extent, private and clinic patients, were located on different floors of the hospital, so that an initial choice as regards these characteristics had always to be made. White clinic patients were most easily available to us and constitute the largest proportion of the control group. Negro and private patients were obtained only in sufficient numbers to test the relationship of these variables to test measures.

Since this was an exploratory as well as a standardization study, during which we not only were in the process of developing the techniques to be reported but also of eliminating others, not all the tests reached their final form at the same time. As soon as a method was in final form, we administered it to a sufficiently large number of normal *Ss* to provide standardization data on reliability and age changes, and it was subsequently administered only to traumatized *Ss* and those on whom oxygen tests had been made. Thus all tests were not administered to all *Ss* or to an equal number of *Ss*. Table 2 shows the distribution of socioeconomic status, sex, and race in the normal and traumatized *Ss* who were given each test, as well as in the total groups. Table 3 gives the age distribution for initial tests and retests.

Procedure

The *Ss* were examined in a hospital room maintained in the same manner as the regular nursery. No soundproofing was available in this room. Whenever extraneous noises were sufficiently loud to startle an *S*, test procedures were repeated. Examinations were carried out between 10:15 A.M. and 3:00 P.M. with

TABLE 2

SOCIOECONOMIC STATUS, RACE, AND SEX OF NORMAL AND TRAUMATIZED SUBJECTS

Test	Sex		Socioeconomic Status		Race		Total *Ss*
	Male	Female	Private	Clinic	Negro	White	
Pain:							
Normal	39	57	33	63	12	84	96
Traumatized	34	21	29	26	17	38	55
Maturation Scale:							
Normal	36	36	19	53	22	50	72
Traumatized	22	6	16	12	7	21	28
Vision Scale:							
Normal	40	38	28	50	16	62	78
Traumatized	23	14	20	17	10	27	37
Irritability:							
Normal	48	54	29	73	22	80	102
Traumatized	18	11	12	17	10	19	29
Tension-							
Normal	47	56	29	74	22	81	103
Traumatized	17	12	11	18	11	18	29
Total *Ss*							
Normal	80	96	54	122	32	144	176
Traumatized	55	26	45	36	23	58	81

TABLE 3

AGE DISTRIBUTION OF INITIAL TESTS (I) AND RETESTS (R) GIVEN TO NORMAL AND TRAUMATIZED SUBJECTS

Test	Day 1*		Day 2		Day 3		Day 4		Day 5†		Total	
	I	R	I	R	I	R	I	R	I	R	I	R
Pain:												
Normal	61	8	9	11	8	9	10	8	8	9	96	45
Traumatized	35	1	6	11	6	13	4	7	4	14	55	46
Maturation Scale:												
Normal	24		14	7	21	4	5	17	8	2	72	30
Traumatized	11		4	1	6	5	4	2	3	5	28	13
Vision Scale:												
Normal	20		17	5	21	4	6	8	14	3	78	20
Traumatized	15		8	2	5	4	1	2	8	15	37	23
Irritability:												
Normal	31		15	6	27	5	13	12	16	5	102	28
Traumatized	16		3	2	3	5	3	3	4	3	29	13
Tension:												
Normal	31		15	6	27	5	13	12	17	5	103	28
Traumatized	17		2	2	3	4	3	2	4	4	29	12

* Includes *Ss* from 12 to 36 hours old with the exception that 15 normal and 3 traumatized *Ss* were given pain tests when less than 12 hours old.

† Includes traumatized *Ss* from 5 to 7 days old.

most *Ss* seen during the morning. As pointed out previously, the number and kind of tests given varied during the course of the study. For those *Ss* who were tested with the final battery, the pain threshold was obtained first. Vision tests were given whenever the infant opened his eyes, and the maturation and tension scale items were given in whatever order best maintained the infant in a satisfactory state. Irritability was rated at the end of the examination. Detailed instructions for each of the procedures is given in the preceding paper (7).

Results[4]

Reliability

In the present research, we are interested in three levels of measurement which require different methods of estimating reliability. In the first place, we are interested in differentiating groups. The reliability of the group measurements would be satisfactory if the error variance of the group means were significantly less than the differences between normal and traumatized groups. This will be dealt with in a later section.

Secondly and primarily, we are concerned with the accuracy of identifying individuals as members of a group. Satisfactory reliability for this purpose would be achieved if errors in measurement were not large enough to change the individual's classification as normal or abnormal.

Thirdly, we are interested in the reliability of individual raw scores, partly because there are standard methods of presenting reliability data in this form and partly because, even where the interest is centered on classifying in two categories, there is always the possibility that raw scores or transformations of them can later be shown to measure degrees of abnormality or normality. We will, therefore, present reliability measures in terms of raw scores, although raw-score reliability will underestimate the reliability of classifying as normal or abnormal. It is only for this twofold classification that validation has been attempted. Three estimates of reliability have been made: (*a*) single-session reliability, given by split-half product-moment correlations, (*b*) test-retest agreement after 24 hours, and (*c*) interscorer agreement.

1. *Split-half reliability.* Only the pain threshold score could be satisfactorily divided into two comparable halves. Split-half correlations (1st and 2nd halves) for normal and traumatized *Ss* are shown in Table 4. The correlations range from .82 to .97. Reliability is highest for the traumatized group, but this is undoubtedly due to the wider range of thresholds obtained in this group. When correlations are calculated separately for traumatized *Ss* with thresholds in the same range as the normal group, the difference in reliability disappears. Age heterogeneity apparently contributes little to raising the correlation. When age is partialled out, the correlation drops only from .87 to .82. It should

TABLE 4

Split-Half Reliability of the Pain Threshold

Subjects	*N*	*r* (corrected)
Normal *Ss*	96	.87
Normal *Ss* (partial *r* without age covariance)	96	.82
Traumatized *Ss* (total)	55	.97
with normal scores	35	.82
with above normal scores	20	.93

[4] The authors are grateful to Robert C. Bilger and to John C. Glidewell for advice on statistical treatment.

be pointed out that momentary fluctuations in the state of an infant are much more marked than in an adult and a split-half reliability estimate includes correspondingly more of the possible variable error.

2. *Test-retest reliability.* Test-retest correlations are reported only for normal *Ss*. Since age is a variable rather than a source of error in the present study, changes associated with it should not be included in a measure of unreliability, unless the effect of age is relatively the same for all *Ss*. The effect of age is undoubtedly *not* the same for all infants in the traumatized group, as they are recovering at different rates from different kinds and degrees of trauma and there is no way of separating the differential changes due to age from those due to unreliability. Therefore test-retest correlations in this group could give little information on reliability. As far as normal *Ss* are concerned, we do not know whether they change differentially with age, although recovery from even a normal birth probably varies. However, there can be no objection to regarding test-retest correlation as at least a minimum estimate of reliability.

Table 5 shows the number of *Ss* who were re-examined on each procedure and the agreement of test and retest score. Product-moment correlations were calculated for Pain Threshold and for Maturation and Vision Scale scores. Percentage of perfect agreement was used in the case of the rating scales. Retest reliability is lower than split-half, as would be expected, but it is satisfactory for the kind of tests used, especially in view of the age changes which are significant at this time of life.

3. *Interscorer reliability.* On all measures except the Pain Threshold, scores depend to some extent upon judgments made by the examiner. In order to determine whether such judgments constituted a major source of unreliability, two examiners simultaneously scored a sample of *Ss*, drawn at intervals throughout the course of the study, to insure that agreement did not vary over a period of time. The examiner who administered the tests had no knowledge of whether *Ss* were traumatized or not traumatized. Table 6 shows the number of *Ss* in each group observed by two examiners and the satisfactorily high interscorer agreement. As in the case of test-retest agreement, Tension ratings appear more reliable than those for Irritability.

Objectivity

Ideally, in such a study as this, all measurements should be made without knowledge of how the *S* is classified. It was not possible to do this routinely, but there are several lines of evidence which

TABLE 5

TEST-RETEST AGREEMENT AFTER TWENTY-FOUR HOURS

Test	Normal *Ss*	Agreement
Pain	24	$r = .69$
Maturation Scale	20	$r = .62$
Vision Scale	20	$r = .62$
Irritability	28	75% perfect agreement
Tension	28	86% perfect agreement

TABLE 6

INTERSCORER AGREEMENT

Test	Normal	Traumatized	Agreement
Maturation Scale	19	5	$r = .97$
Vision Scale	17	4	$r = .90$
Irritability	20	5	68% perfect agreement
Tension	20	4	79% perfect agreement

indicate that we were successful in avoiding bias.

1. While there was often knowledge of the presence or absence of trauma, whether it was mild, moderate, or severe was *not* known. As will be subsequently shown, test results not only differentiated normal from traumatized *Ss* but also were correlated with the degree of trauma.

2. Interscorer reliability was high (see Table 6) for *Ss* scored by two examiners simultaneously. The examiner who administered the tests on these check cases was without knowledge of the *S*'s classification.

3. A group of 16 traumatized and 31 normal *Ss* was examined without any knowledge of classification. The 16 traumatized *Ss* had been given a total of 41 tests and the 31 normal *Ss* a total of 68 tests. The scores of each *S* on the tests he was given were paired with the scores of an *S* tested with knowledge of classification. Pairing took into account age, race, and kind and degree of trauma (when present). The Sign Test was used to test the hypothesis that the score differences obtained under the two conditions were randomly distributed as to direction. For this analysis, scores on all tests were considered simultaneously. Results showed that there were no significant differences in either the normal or traumatized *Ss* to suggest that scores were more "normal" or more "abnormal" when there was knowledge of the classification of the infant than when there was not such knowledge. Since bias might occur more easily on some of the more subjective tests than on others, *F* tests were applied to each test separately. Of the 10 comparisons made, normal and traumatized groups being treated separately, only one indicated a difference significant at the .05 level. That was on the Tension Rating scale, where 14 normal *Ss* tested with knowledge were rated lower than those tested without knowledge.

Variables Affecting the Performance

The normal and traumatized groups differ considerably in respect to age, sex, race, and private or clinic status, as may be seen from Tables 2 and 3. Before the performance of the two groups could be compared, therefore, it was necessary to determine whether or not these characteristics were related to the scores obtained by normal *Ss*. The existence of a relationship in traumatized *Ss* which did not exist in normal *Ss* would not confound the results but would be assumed to be an interaction effect dependent upon trauma.

In order to determine the effect of so many variables without allowing concomitant variation, it was necessary to select smaller samples from the pool of *Ss* in such a way that factors other than the one being tested were held constant. Male and female *Ss*, paired for race, socioeconomic status, and age, did not show differences on any of the tests. Private and clinic *Ss*, paired for race and age but with sex allowed to vary since it did not influence performance, also did not differ from one another on any of the tests. The effects of race, age, retesting and condition of the *Ss* were similarly tested. The *F* ratio was used to compare scores on the Pain Threshold test and on the Maturation and Vision Scales. Chi square was employed in comparing scores on the two rating scales. The ratings were categorized as "o" or "not o" in order to obtain sufficiently large theoretical frequencies but, even so, the expected frequencies were less than 5 in two of the eight comparisons. However, none of the probabilities calculated from the chi-square distribution approached significance and would not do so even if calculated directly.

Private-clinic status and sex of the *S* did not measurably affect performance on any of our tests, and there was no practice or learning effect from retesting when age was held constant. However, age itself was significantly related to score on three tests. Older *Ss* were found to be more sensitive than younger to Pain, and to perform better on the Maturation and Vision Scales. Negro *Ss* were also superior to whites on the Maturation and Vision Scales but there was no difference on the other tests. Unless the condition of *Ss* being given the Maturation Scale was "satisfactory," as previously defined, significantly poorer scores were obtained. This factor was not important, in so far as it was measured, on other tests. Where we have made assertions that a factor did affect performance, the usual 5 or 1 per cent levels of significance of difference are to be understood. Where the statement is made that a factor did *not* affect performance, the *F* ratios were in every case smaller than 1.0, and the highest chi-square probability was .18. There was, therefore, nothing to suggest that an effect would be demonstrated if another sample or a larger one were obtained.

Differences Between Normal and Traumatized Groups

Mean difference. In order to compare the performance of normal and traumatized groups on the test battery,

it was necessary to equate the groups for those variables which have been shown to be related to test score. As neither the irritability nor tension ratings were affected by the variables tested, no pairing was necessary on these measures. The entire sample of normal and traumatized *Ss* could be used, therefore, with the exception of those rated on irritability a day or two after circumcision. Statistical analysis, not reported, showed that this factor did raise the irritability rating. Table 7 shows the size of the groups, the variables on which they were equated by pairing, the mean scores of the two groups, the statistic used in estimating probabilities, and the probability that differences between groups are due to chance. The means are included on all five measures, although they were of course not used when chi square was the comparison statistic. On both the pain thresholds and the Vision Scale there was a significant difference in the variance of the traumatized and normal groups. In evaluating the mean difference, therefore, t was calculated according to the Cochran-Cox method with no assumptions about variance. On all tests the performance of the traumatized groups was significantly poorer than that of the normal groups.

Shape of the distributions. In the preceding section, evidence was presented that the normal and traumatized groups differed significantly, but nothing was said about the shape of the distributions. Figure 1 shows the distribution of pain thresholds in the two groups. Percentage frequency rather than frequency is plotted on the ordinate to equate the size of the groups. Standard scores are plotted on the abscissa so that thresholds of individuals tested on different days could be combined. Since the standard score transformation is a linear transformation, the shape of the curve is altered only to the extent that the distributions of the several days show differences in skewness. Standard scores of traumatized *Ss* are based on the normal group.

TABLE 7

COMPARISON OF NORMAL AND TRAUMATIZED GROUPS ON FIVE TESTS

Test	*N*	Variables Controlled	Mean Scores	Comparison Statistic	*p*
Pain Threshold					
Normal	55	Age	165	*t* test	.01
Traumatized	55		270		
Maturation Scale					
Normal	28	Age, race	13.0	*F* test	.05
Traumatized	28		10.6		
Vision Scale					
Normal	37	Age, race	6.8	*t* test	.01
Traumatized	37		4.2		
Irritability					
Normal	91	———	.12	Chi square	.01
Traumatized	29		.61		
Tension					
Normal	103	———	.08	Chi square	.01
Traumatized	29		.48		

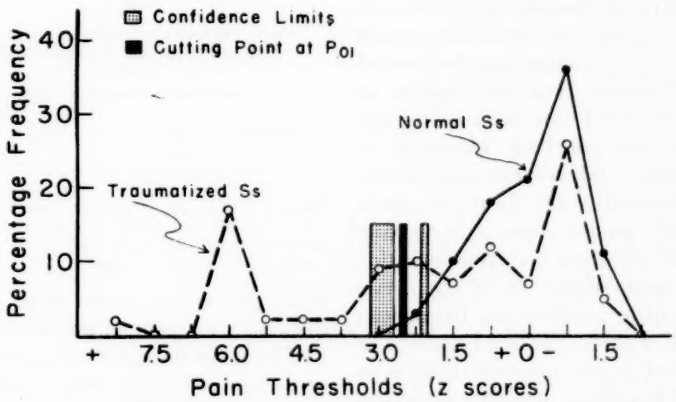


FIG. 1. Pain thresholds (in standard score form) of normal and traumatized subjects.

The graph shows a skewed distribution in the normal group, the skewness probably reflecting the effect of a lower limit to the pain threshold. Thresholds of the traumatized group, on the other hand, cover a much wider range and do not appear to fall into any type of unimodal distribution. Unless it is assumed that the considerable variability has distorted what would otherwise be a unimodal distribution, we must assume a bi- or multimodal distribution. Since the tendency to bimodality is also present in each of the three subsamples of mild, moderate, and severely traumatized infants, the hypothesis of bimodality is strengthened.

If thresholds of traumatized *Ss* are not unimodally distributed, what does this signify? It can only mean that pain sensitivity is not affected in a unitary fashion under conditions of trauma. More than one factor must be present. It is possible that the presence of a new factor is due to the greater intensity of stimulation used with some of the traumatized *Ss*, rather than to the trauma per se. This would be the case if other sense modalities were activited when a certain intensity of stimulation is reached. It is tempting, however, to speculate that the new factor is related to brain functioning. Is there a threshold for impairment of brain functioning such that on one side of the threshold, the same factors determining pain sensitivity in normal brains are operating while, once beyond the threshold, sensitivity is determined by changed condition of the brain? It is idle to speculate on the brain physiology which might be involved; there are many physiological phenomena which show this all-or-none character. For present purposes, the point is of interest in selecting cutting scores and making predictions. If the nonunitary character of pain sensitivity is due to changes in brain functioning, we would expect to find that *Ss* who later show evidence of brain damage would be selected from among those who form a second mode and not from among those who fall at the upper end of the normal distribution.

Graphic distributions of the other four tests are not presented. Interpretation of them is complicated by the fact that scoring was empirically determined

on the basis of observation by the authors, and changes in the scoring system would, of course, change the character of the distributions. The distribution of Maturation and Vision Scale scores was similar to that for Pain Threshold in both normal and traumatized groups. On the Irritability and Tension ratings, the normal groups show a heavy concentration of scores receiving 0 rating with a rapid falling off of the curve. There is little tendency for frequencies to pile up at the tail as in a J curve. The traumatized groups, however, do show an increased frequency of higher ratings as well as a wider range.

Cutting Points and Normative Data

In order to identify those *Ss* among whom we expect to find later evidence of brain damage, it is desirable to establish a cutting point. It would be possible to correlate scores on each of the tests with later results. However, since we are dealing with a phenomenon of low frequency in the total population, it is likely that the percentage of "hits" can be increased by restricting ourselves to a more eligible subsample. If the bimodal distribution of scores in the traumatized group is related to brain function, it becomes even more important to separate the population forming the second mode from that forming the first mode, since there would be no reason to expect differences in scores around the normal mode to be correlated with later damage.

It is also desirable to have a cutting point which will include a minimum of false positives, i.e., normal *Ss* incorrectly called abnormal. From the application of Bayes's theorem, which Meehl has made (12), it is easily demonstrated that when a clinical group occurs with low frequency in the general population, there will be more incorrect than correct predicitions even when the percentage of true positives approaches 100, unless false positives are kept at a minimum.

TABLE 8

PAIN THRESHOLDS IN VOLTS OF NORMAL SUBJECTS FOR THE FIRST FIVE DAYS OF LIFE

Statistic	Day					
	.25	1	2	3	4	5
N	15	54	20	17	18	17
Mean	205	185	155	140	120	98
SD	70	68	50	55	35	35
$p = .01$	380	350	280	270	195	183

These considerations led us to select a cutting point at the extreme of the normal distribution—that point below which only 1 per cent of the normal population would fall. Cutting points for pain thresholds, Maturation, and Vision scores were set at t values with probabilities at the .02 level (or .01 level for a single tail), rather than at the observed $p = .01$ value in the sample. Basing cutting points on t values, when justified by an approximately normal distribution, appears preferable to using sample percentile points. Separate cutting points were determined for each day, and, in the case of Maturation and Vision scores, for both Negroes and whites. These cutting points as well as the means and standard deviation are shown in Tables 8, 9, and 10.[5] The reliability of the cutting points for pain thresholds can be estimated from the 95 per cent confidence limits, which are shown in Fig. 1. Bars rather than lines

[5] Because race and age did affect performance they required separate norms. An additional 89 *Ss* were therefore tested in order to increase the size of *N* on the Maturation, Vision, Irritability, and Tension scales.

TABLE 9

MATURATION SCALE SCORES OF NORMAL SUBJECTS FOR THE FIRST FIVE DAYS OF LIFE

Statistic	Day				
	1	2	3	4	5
White *Ss*:					
N	37	21	20	21	20
Mean	11.2	11.9	12.8	13.5	14.1
SD	4.3	2.8	2.3	2.9	2.2
$p = .01$	.8	3.9	7.1	6.3	8.6
Negro *Ss*:					
N	28		20		
Mean	12.4		13.7		
SD	3.0		2.6		
$p = .01$	4.9		7.0		

identify the upper and lower confidence limits since this is a composite curve of the samples on each of the five days. The limits are sufficiently narrow so that false positives would vary only between 0 and 3 per cent if the cut is located at any point within the interval. The number of true positives could vary more widely, but, even at the upper limits, none of the cases at the second mode is excluded. The confidence intervals for cutting points on the Vision Scale are also narrow. False positives would vary only from 0 to 2 per cent and true positives from 38 to 43 per cent. The Maturation Scale shows less discrimination than the other measures to begin with. While changes in the location of the cut would not increase the number of false positives, the discrimination of true positives could be cut from 25 to 7 per cent.

TABLE 10

VISION SCORES OF NORMAL SUBJECTS FOR THE FIRST FIVE DAYS OF LIFE

Statistic	Day				
	1	2	3	4	5
White *Ss*:					
N	37	26	23	20	24
Mean	4.9	6.5	6.4	6.1	6.9
SD	2.2	1.3	1.4	1.5	1.9
$p = .01$	0	3.4	2.8	2.3	2.3
Negro *Ss*:					
N	23		21		
Mean	5.4		7.6		
SD	2.0		2.0		
$p = .01$	.5		2.6		

In determining cutting points for the Irritability and Tension Scales, it was necessary to use the observed p values since the distributions depart radically from the normal curve as can be seen from Table 11. However, ties in score (at a rating of 1.0) occurred in the lower 6 per cent of Irritability Ratings and 3 per cent of Tension Ratings so that a value unique to the first percentile could not be determined. While it is possible to divide the tied cases into appropriate proportions above and below the cutting point, this is not satisfactory when a decision must be made about classifying an individual *S*. The cut must be placed either just below or just above the score. The decision was made to place it above the rating of 1.0 on the Irritability Scale, thus excluding 99.5 per cent of normal *Ss*, and below the 1.0 rating on the Tension Scale, which excluded only 96.8 per cent of normals. These decisions took into account the shape of both the normal and traumatized group distributions and may to some extent have capitalized on chance fluctuations. Estimates

TABLE 11

PERCENTAGE OF NORMAL SUBJECTS RECEIVING A GIVEN IRRITABILITY OR TENSION RATING SCORE ($N = 186$)

Scale	Rating Score					
	0	.25	.5	.75	1.0	1.25 or above
Irritability	76.3	5.9	10.2	1.1	5.9	0.5
Tension	79.1	8.6	7.5	1.1	2.7	0.5

of the reliability of these cuts are necessarily crude. If 95 per cent confidence limits based on the binomial distribution are determined, we can expect the percentage of false positives on the Irritability Scale to vary from 0 to 4 per cent and the percentage of false positives on the Tension Scale to vary from 1 to 7 per cent.

These cutting points are only tentatively identified. The optimal cut, in the sense of giving minimal overlap in both directions, lies at the intersection of the normal and traumatized group distributions. For the reasons outlined above, we did not feel that such a cut would be useful in a situation where base rates are presumably low and where it is more important to exclude false positives than to reduce false negatives. Final decisions about the value of these cutting points depends upon their relationship to signs of brain damage in the *Ss*' subsequent development. For this reason, and because the points are determined entirely by the normal samples, which are reasonably large and reliably measured, we did not reserve a portion of the data for cross validation.

A rough measure of the relative discriminating power of the five tests may be obtained by comparing the percentage of traumatized *Ss* called abnormal. Table 12 shows the percentage of *Ss* who score on the abnormal side of the cutting point on any one or more tests and on each test separately. When *Ss* were retested, the poorest performance has been taken as the score on a test. These data are supplied for the normal and traumatized groups and for the three subsamples of traumatized *Ss*. Pain thresholds and the Vision Scale are superior, but all tests identify some *Ss* as abnormal. The percentage identified as abnormal appears to increase with the degree of trauma and, if scores on all tests are considered, is statistically significant at the .01 level when tested by chi square.

Intercorrelations

How are the various tests related to one another? Since they were designed initially to detect differences between traumatized and normal *Ss* rather than among normals, the intercorrelations might be expected to differ for the two kinds of *Ss*. They are therefore presented separately, as well as for the total group, in Table 13. As pointed out earlier, all tests were not given to all *Ss*. The traumatized cases included in the correlations are all those *Ss* who had

TABLE 12

PERCENTAGE OF SUBJECTS IDENTIFIED AS ABNORMAL BY SCORES BELOW THE CUTTING POINT ON THE DAY OF POOREST PERFORMANCE

Test	Normal	Total Traumatized	Mild Trauma	Moderate Trauma	Severe Trauma
Pain	1	42	43	30	57
Maturation Scale	0	25	(33)*	12	(50)
Vision Scale	1	41	(17)	31	60
Irritability	1	28	(33)	8	46
Tension	3	34	(33)	23	46
Any one or more tests	4	51	46	37	84

*Percentages in parentheses are based on an *N* of less than ten.

TABLE 13
INTERCORRELATIONS AMONG TEST PROCEDURES

Test	Normal *Ss*		Traumatized *Ss*		Total group	
	N	*r*	*N*	*r*	*N*	*r*
Pain-Mat.	46	.14	26	−.04	72	.06
Pain-Vision	46	−.11	40	−.03	86	.06
Pain-Irrit.	46	.43**	42	.08	88	.07
Pain-Tens.	46	.03	41	.16	87	.21*
Mat.-Vision	46	.38**	22	.31	68	.36**
Mat.-Irrit.	46	.09	27	−.02	73	.12
Mat.-Tens.	46	.01	26	.20	72	.20
Vision-Irrit.	46	−.24	30	.43*	76	.49**
Vision-Tens.	46	−.17	30	.15	76	.13
Irrit.-Tens.	46	.21	42	.30*	88	.39**

scores on at least two tests. The 46 normal cases are the total number of *Ss* to whom all five tests were given. Standard scores were used and the direction of scores changed so that all correlations may be similarly interpreted.

Table 13 shows that the intercorrelations are, in fact, different for the two groups. The Maturation and Vision Scales, both presumably measuring developmental level and having considerable spread in both groups, are positively correlated with one another, although the correlation is not statistically significant in the traumatized group. As would be expected, Irritability and Tension are positively correlated in the traumatized group but not in the normal. They are measures designed to detect deviations in the direction of abnormality and therefore have little variability in the normal group. Irritability, a measure of abnormality, and Vision, a measure of developmental level, are significantly related only in the traumatized subjects.

The finding of a correlation among normal *Ss* between Pain, which does have spread within the normal range, and Irritability, which does not, is somewhat surprising. One can only suggest that the irritable infant is responding to many stimuli in a diffuse fashion and therefore tends to be less sensitive or less set for responding to a specific stimulus in a specific way. This same analysis might be expected to hold true among traumatized *Ss* except for the fact that some traumatized infants become more irritable while others become obtunded. The obtunded infant is scored as normal in Irritability but will, of course, be relatively insensitive to pain as well as to other stimuli.

Perhaps the most significant finding is that all intercorrelations are low. This is understandable since trauma to the newborn may be manifested in a variety of ways, some of which are incompatible with one another, as, for example, obtundity and hyperirritability. Such relatively low intercorrelations among the different tests, together with the adequate discriminating ability of each considered separately, points to the advisability of using a combined score on all procedures as an impairment index. This would achieve one of the characteristics desirable in constructing a test battery—i.e., low intercorrelations among the tests and high correlations with the criterion.

Another possible way of improving predictive accuracy would be to consider the length of time that an *S*'s performance remains abnormal. In retesting traumatized *Ss*, we observed that the length of time scores remained abnormal varied considerably from infant to infant. The duration of an abnormal performance thus provides another dimension along which to measure the infant's response to trauma and offers promise as an additional way of identifying those *Ss* on whom trauma will leave a permanent imprint.

Discussion

The group of tests we have used samples much of the repertoire of an infant's response to his environment. All of the responses are relatively simple, but they represent a substantial portion of the most complicated behavior which an infant of this age can show. How complicated is such behavior? With the exception of the two rating scales, the tests can be described as measuring sensorimotor ability, i.e., (*a*) the capacity to respond at all to various kinds of sensory stimuli, and (*b*) the extent to which the response is specific to a particular stimulus. The ratings of irritability and of muscular tension provide two more dimensions along which all responses of an infant, both spontaneous and elicited, may be described. We should like to know whether measuring such behavior gives any information about either past or future development.

Does sensorimotor functioning reflect the developmental level of the nervous system? Gesell (6) has carried out extensive studies of premature and full term newborns which suggest that simple sensorimotor abilities are in the process of development during the last months of fetal life and therefore should provide measures of the stage of development of the nervous system. This pioneer work needs to be confirmed, however, by studies using standardized procedures whose reliability can be established. The difficulty we experienced in adapting from Gesell a reliable maturational scale and the widely divergent placement of similar items on various infant scales indicate that slight variations in method can produce large differences in results.

Does sensorimotor functioning at birth predict later development? Efforts to answer this question have extended over several decades, but it is still not possible to formulate definite conclusions. This is partly because similar data have been interpreted by different standards. A correlation of a given size does not arouse the same response in all psychologists. The data, however, are also conflicting. Some studies have shown small negative correlations between performance in early infancy and several years later (10, p. 637f). On the other hand, sizable positive correlations have been reported between tests at six months and tests as late as three years (14). Such factors as inadequately standardized tests, difficulties in controlling the infant's momentary state (3), differences in the kind of infant tests, and differences in the kind of development which is being predicted all contribute to the confusion.

Irwin (9) has questioned the reasonableness of expecting to predict "intelligent" behavior from neonatal response under even the most favorable conditions and with sensitive tests. Neurophysiological work suggests that the newborn is not capable of cortical functioning (4, p. 60f). While there is opposition to this view, the argument is mainly whether there is no cortical activity or some minimal amount (11). Irwin's (9) position is essentially that you cannot expect to predict the future functioning of the cerebral cortex on the basis of tests made at a time when the cortex is *nonfunctional*. On logical grounds alone, this is not a necessary conclusion. There is no a priori reason why the functioning of subcortical structures should be uncorrelated with later cortical development. From Hebb's (8, p. 109ff) interesting hypothesis that the higher organism is the slower learner initially, one might even infer that there is a negative correlation. Practically, however, the greater the gap between phenomena observed and those to be predicted, the more difficult it is likely to be to establish relationships.

The present work is not primarily concerned with predicting the relative superiority of "normal" individuals, but rather in determining whether external trauma has caused brain injury. We did find that a considerable percentage of traumatized infants show impaired functioning as compared with nontraumatized newborns, and that such impairment is related to clinical judgments of severity of trauma. But

will measures of impairment of a newborn predict the extent of later impairment? The question cannot be answered at the present time. One can say only that it seems reasonable to assume that the greater the present trauma, the greater the likelihood that some cells will suffer irreversible damage. Since the cerebral cortex of the infant is relatively nonfunctional, is it likely that trauma to the cortex could be detected by newborn tests? The answer would seem to be no, if the cortex is, in fact, completely nonfunctional and if the trauma were limited only to the cortex. Destruction of a part of the nervous system before it is mature enough to function will not give rise to the expected symptoms until "the time arrives for that structure to play its proper role" (4, p. 69). However, many kinds of trauma, and especially anoxia, may produce diffuse, multiple lesions in both cortical and subcortical structures (1). With such trauma, the extent of disruption of subcortical functions might provide an index of the degree of total damage.

We suggested that a general impairment index could be computed by combining the number of tests showing impairment and the length of time impairment persists. The extent to which such an index will be valuable depends on whether or not specific areas of damage are associated with impairment on specific tests. Without discarding the possibility of specificity entirely, we would relegate it to a minor role. Functioning of a newborn is relatively undifferentiated and it seems likely that the extent of trauma would be the most important factor in determining subsequent impairment. This could be true either if sizable injuries in any area affect most functions, or if there is a tendency for the kinds of trauma to which a newborn is exposed to cause injury in the same areas because these are more susceptible to insult.

It would be interesting to know more about the nature of the impairment that occurs. Because the intensity of the pain stimulus was systematically varied, we were able to observe that in infants whose pain thresholds were abnormally high, the ability to perceive the stimulus did *not* seem impaired, since many *Ss* cried or gave other general responses to weak stimuli. Similarly, motor pathways were apparently intact, since the required leg movements could occur independently of stimulation. The difficulty, then, must lie in a failure to integrate stimulus and response in such a way that a response appropriate or specific to the stimulus could occur. In the development of the nervous system, sensory and motor nerves are in contact with their respective organs before it is possible for an excitation to pass from sensory to motor mechanism. We may speculate that, in the loss of the specific connection between stimulus and response under conditions of trauma, we see an example of the phenomenon observed in adult brain-damaged individuals, namely, that those abilities most recently acquired are most readily lost.

One other phenomenon deserves mention. Some of the traumatized infants might be described as hyperreactive with increased muscular tension and irritability. These infants were sensitive to any kind of mild stimulation but gave generalized rather than specific responses. Other traumatized infants showed diminished general activity, were flaccid and apathetic. Deviations in either direction could impair performance on various tests. The direction of deviation was not taken directly into account in the scoring, although apathetic infants were more likely to be penalized on the Maturation Scale. Are these two kinds of impairment related to the severity of the trauma? Is it possible that they are analogous to stages in the development of coma in adults? Initially, "in anesthesia and in coma induced by changes in the internal environment (decrease in the oxygen tension or in the blood sugar level) the cortex and brain stem suffer opposite changes in excitability" (5, p. 212). While cortical activity is diminished, there is increased reactivity of the brain stem. As coma deepens, brain-stem activity likewise decreases (5, p. 217).

We have been interested in differentiating those infants whose trauma is contemporaneous with the examination. What can be said of the possibility of detecting injury which has occurred some time prior to birth? This will depend on the extent to which our measures reflect temporary effects due to a present disturbance of the nervous system. The newborn traumatized early in intrauterine life should not appear irritable or obtunded, and if the damage suffered has been restricted to the cortex there might be no behavioral manifestations at birth. On the other hand, if subcortical centers have been damaged, we might expect impairment of sensorimotor functioning. We had the opportunity of testing three mongols—one of the few kinds of fetal injury which, because of the physical peculiarities, can be detected at birth. All three did show impairment on Vision and Maturation Scales and also showed the marked muscular flaccidity which characterizes this condition. Their pain thresholds, however, were normal and, except for one infant with an intestinal infection, they were not irritable.

In concluding, an encouraging word ought to be said for the value of infants as psychological *Ss*. In the last decade, they have been largely

overlooked, yet their very youngness offers unique advantages in psychological areas where the complexity of social and other environmental influences makes it difficult to disentangle the relevant variables. There is ample room for improvement in techniques of measurement. Our relatively crude efforts at objectification have been rewarding and have suggested many lines which might profitably be followed with more precise methods.

We aimed to develop a short battery of tests requiring minimal equipment in order that a large number of *Ss* could be studied quickly. Because of these requirements, some promising methods were dropped which ought to be investigated more fully. With photographic equipment, for example, thresholds could be obtained for the pupillary response to light. Electronic equipment is now available that would permit measurement of several aspects of infant activity level. Early work by Richter (15) and Wenger (17) related the height of the skin resistance level to states of tension and wakefulness. Individual differences in hydration and in toughness of the skin present serious problems, but the desirability of obtaining an independent and objective measure of the state of an infant is great.

The most conspicuous omission from our battery is a measure of learning. A series of early studies attempting to establish classical conditioning in the newborn were disappointing (13, p. 376f) but these studies did not exhaust the possibilities of studying the conditions for and kinds of changes in behavior which newborns can exhibit. We found that, except for the rating scales, performance on our tests improved with age during the first five days of life. In the case of pain thresholds, the improvement was linearly related to age change, as can be seen in Figure 2. Is this improvement due to maturation, to experience, to recovery from the trauma of birth, or to a combination of these factors?

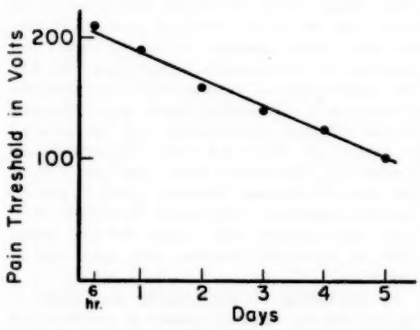


FIG. 2. Mean pain thresholds of normal subjects on the first five days of life.

Without raising the question of whether the improvement is due to maturation or learning, one may ask what it is that has improved. In the case of the pain thresholds, for example, has the infant become more sensitive to pain as he grows older, or does the age change represent increased ability to give a differentiated response? Our results cannot separate the two possibilities. It would be interesting to determine whether the age changes are still present when any kind of response to the stimulus is recorded, i.e., when a generalized rather than a differentiated response is measured. If there are *no* age changes in the ability to make a generalized response, then the hypothesis of increased sensitivity could be rejected. This procedure was actually used in two previous studies, one by Dockeray (2) in 1934, and one by Sherman, Sherman, and Flory (16) in 1936. Unfortunately, their results disagree, which is perhaps to be expected since only very crude control of stimulus intensity was possible at that time. Repetition of this experiment with modern methods of stimulus control should be able to provide a more conclusive answer.

SUMMARY

Five test procedures, described in the preceding paper (7), were administered to 265 infants without prenatal, perinatal, or postnatal complications and to 81 infants suffering from anoxia, mechanical birth injury, or diseases or infections associated with brain damage. The traumatized newborns composed nearly the total population of such infants born at the St. Louis Maternity Hospital during a two-year period.

The five tests consist of a Pain Threshold Test, a Maturation Scale, a Vision Scale, an Irritability Rating and a Muscle Tension Rating. Reliability of the procedures was variously measured by split-half correlation, test-retest agreement, and interscorer agreement as applicable. All of the tests appeared to be satisfactorily reliable. A sample of 109 test scores obtained without knowledge of *Ss'* classification did not differ significantly from those obtained under the usual conditions of partial knowledge.

Norms have been presented for each test, with separate norms provided for each of the first five days of life and for Negro and white *Ss* where these variables were related to performance. Older *Ss* were found to be more sensitive than younger on the Pain Threshold Test and to perform better on the Maturation and Vision Scales. Negro *Ss* were superior to whites on both the Maturation and Vision Scales, but there was no race difference on the other tests. Private-clinic status and sex of the *S* did not measurably affect performance. There was no practice or learning effect from retesting when age was held constant.

Normal and traumatized groups, paired for relevant variables, obtained significantly different scores on all tests. When a cutting point at the poorer extreme of the normal distribution was selected, all tests identified some traumatized *Ss* as abnormal while false positives ranged only from 1 to 3 per cent. The percentage identified as abnormal increased with the seriousness, as rated by pediatric judges, of the trauma. Since intercorrelations of the five tests were low, a combined abnormality score was tentatively recommended.

The question of whether newborn behavior can predict either past or future development of the infant was discussed. Cautious consideration was also given to the relationship between the present findings and neuro-physiological knowledge about functioning of the newborn brain.

REFERENCES

1. COURVILLE, C. B. *Cerebral anoxia.* Los Angeles: San Lucas, 1953.
2. DOCKERAY, F. C., & RICE, CHARLOTTE. Response of newborn infants to pain stimulation. *Ohio State Univer. Studies. Contrib. Psychol.*, 1934, No. 12, 82-93.
3. ESCALONA, SIBYLLE. The use of infant tests for predictive purposes. *Bull. Menninger Clin.*, 1950, **14**, 117-128.
4. FORD, F. R. *Diseases of the nervous system in infancy, childhood and adolescence.* (3rd Ed.) Springfield, Ill.: Charles C Thomas, 1952.
5. GELLHORN, E. *Physiological foundations of neurology and psychiatry.* Minneapolis: Univer. of Minnesota Press, 1953.
6. GESELL, A., & ARMATRUDA, CATHERINE. *Developmental diagnosis.* New York: Hoeber, 1941.
7. GRAHAM, FRANCES K. Behavioral differences beween normal and traumatized newborns. I. The test procedures. *Psychol. Monogr.*, 1956, **70**, No. 20 (Whole No. 427).
8. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
9. IRWIN, O. C. Can infants have IQ's. *Psychol. Rev.*, 1942, **49**, 69-79.
10. JONES, H. E. The environment and mental development. In L. Carmichael (Ed.) *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954.
11. KLEITMAN, N. The role of the cerebral cortex in the development and maintenance of consciousness. In H. A. Abramson (Ed.) *Problems of consciousness.* New York: Josiah Macy Jr. Foundation, 1955.
12. MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194-216.
13. MUNN, N. L. Learning in children. In L. Carmichael (Ed.) *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954.
14. NELSON, VIRGINIA L., & RICHARDS, T. W. Studies in mental development. I. Performance on Gesell items at six months and its predictive value for performance on mental tests at two and three years. *J. genet. Psychol.*, 1938, **52**, 303-325.
15. RICHTER, C. High electrical resistance of the skin of newborn infants and its significance. *Amer. J. Dis. Children*, 1930, **40**, 18-26.
16. SHERMAN, M., SHERMAN, IRENE, & FLORY, C. D. Infant behavior. *Comp. Psychol. Monogr.*, 1936, **12**, No. 4.
17. WENGER, M. A., & IRWIN, O. C. Fluctuations in skin resistance of infants and adults and their relation to muscular processes. *Univer. Iowa Stud. Child Welf.*, 1936, **12**, 141-179.

(Accepted for publication December 19, 1955)

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN